NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

DEVELOPMENT OF THE NONTECHNICAL SKILLS FOR OFFICERS OF THE DECK (NTSOD) RATING FORM

by

W. Max Long

December 2010

Thesis Advisor: Paul E. O'Connor
Second Reader: Michael E. McCauley

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 2010	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Development of the Nontechnical Skills for Officers of the Deck (NTSOD) Rating Form			**5. FUNDING NUMBERS**
6. AUTHOR W. Max Long			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number NPS-2010-0060.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This thesis describes the need for and development of the Nontechnical Skills for Officers of the Deck rating form. The purpose of the NTSOD rating system is to provide a framework for evaluating the nontechnical skills of U.S. Navy Surface Warfare Officers of the Deck (OOD). Nontechnical skills are the cognitive, social, and personal skills that complement technical skills. The NTSOD system supplements the current OOD qualification process by providing an objective and documented assessment of the nontechnical skills of OOD candidates. The system consists of four categories of behavior (leadership, communication, situational awareness, and decision making), which are subdivided into 10 more specific, and observable, elements of behavior. The framework was developed through the analysis of data collected from qualified OODs. When properly utilized, the authors believe that the NTSOD system can improve the skills of OODs, and increase the overall mission effectiveness and safety of the surface fleet by providing feedback on a crucial aspect of OOD performance that is not consistently evaluated across the Surface Fleet.

14. SUBJECT TERMS Surface Warfare Officers, Officer of the Deck, Behavioral Markers, Behavioral Marker Systems, Nontechnical Skills, NTSOD			**15. NUMBER OF PAGES** 81
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

DEVELOPMENT OF THE NONTECHNICAL SKILLS FOR OFFICERS OF THE DECK (NTSOD) RATING FORM

W. Max Long
Lieutenant, United States Navy
B.A., University of Kansas, 2004

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN HUMAN SYSTEMS INTEGRATION

from the

NAVAL POSTGRADUATE SCHOOL
December 2010

Author: W. Max Long

Approved by: Paul E. O'Connor
Thesis Advisor

Michael E. McCauley
Second Reader

Robert F. Dell
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis describes the need for and development of the Nontechnical Skills for Officers of the Deck rating form. The purpose of the NTSOD rating system is to provide a framework for evaluating the nontechnical skills of U.S. Navy Surface Warfare Officers of the Deck (OOD). Nontechnical skills are the cognitive, social, and personal skills that complement technical skills. The NTSOD system supplements the current OOD qualification process by providing an objective and documented assessment of the nontechnical skills of OOD candidates. The system consists of four categories of behavior (leadership, communication, situational awareness, and decision making), which are subdivided into 10 more specific, and observable, elements of behavior. The framework was developed through the analysis of data collected from qualified OODs. When properly utilized, the authors believe that the NTSOD system can improve the skills of OODs, and increase the overall mission effectiveness and safety of the surface fleet by providing feedback on a crucial aspect of OOD performance that is not consistently evaluated across the Surface Fleet.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ANTS	Anesthetists’ Nontechnical Skills
AWACS	Air Force Airborne Warning and Control System
BMOW	Boatswain’s Mate of the Watch
CDO	Command Duty Officer
CIC	Combat Information Center
CICWO	Combat Information Center Watch Officer
CIT	Critical Incident Technique
CO	Commanding Officer
Comms	Communications
Conn	Conning Officer
CRM	Crew Resource Management
DMSC	Dynamic Model of Situated Cognition
EOOW	Engineering Officer of the Watch
FPO	Force Protection Officer
GPS	Global Positioning System
HFACS	Human Factors Analysis and Classification System
HRO	High Reliability Organizations
HSI	Human Systems Integration
JOOD	Junior Officer of the Deck
JOOW	Junior Officer of the Watch
κ	Cohen’s Kappa
MOBI	Man Overboard Indicator
MOBOARD	Maneuvering Board
M/V	Motor Vessel
OPLAT	Oil Platform
NOTECHS	Pilots’ Nontechnical Skills Behavioral Marker System
NOTSS	Nontechnical Skills for Surgeons
NTSOD	Nontechnical Skills for Officer’s of the Deck
OOD	Officer of the Deck
PC	Patrol Craft

PQS	Personnel Qualification Standards
QM	Quartermaster
QMOW	Quartermaster of the Watch
SME	Subject Matter Expert
SWO	Surface Warfare Officer
TAO	Tactical Action Officer
UNREP	Underway Replenishment
USCGC	United States Coast Guard Cutter
VERTREP	Vertical Replenishment
XO	Executive Officer

EXECUTIVE SUMMARY

Research has shown that greater than 80% of incidents in high-risk organizations result from human error. The maritime industry is no exception. It is estimated that 75% to 96% of maritime casualties are caused by human error. Although training and feedback are provided on the technical skills and knowledge of operators, nontechnical skills are rarely addressed. Nontechnical skills are the cognitive, social, and personal resource skills that complement technical skills and contribute to safe and efficient task performance. Several high-risk industries such as medicine, nuclear power plants, and civilian aviation have created behavioral marker systems that provide a mechanism for observing and assessing the nontechnical skills of operators. However, there is currently no system available to evaluate nontechnical skills of Surface Warfare Officers.

It is not possible to create a system that is capable of assessing the nontechnical skills required for every watch station on a ship. There are simply too many individuals with too wide a range of responsibilities. However, as the Officer of the Deck is the most important watch station on surface ships, it was the most logical place to begin addressing nontechnical skills.

This thesis describes the development of the Nontechnical Skills for Officers of the Deck (NTSOD) rating form. The purpose of the NTSOD rating system is to provide a framework for evaluating the nontechnical skills of U.S. Navy Surface Warfare OODs. The NTSOD system supplements the current OOD qualification process by providing an objective and documented assessment of the nontechnical skills of OOD candidates.

The NTSOD system consists of four categories of behavior (leadership, communication, situational awareness, and decision making), which are subdivided into 10 more specific, and observable, elements of behavior. The framework was developed through the analysis of data collected from qualified OODs. When properly utilized, the authors believe that the NTSOD system can produce skilled OODs, and increase the

overall mission effectiveness and safety of the surface fleet by providing feedback on a crucial aspect of OOD performance that is currently not consistently evaluated across the Surface Fleet.

ACKNOWLEDGMENTS

First, I would like to thank Dr. Paul O'Connor for advising me on this thesis and acting as an additional rater during the analysis. While this thesis and the associated documents traveled in directions that you probably did not foresee, the original idea was yours. Some portions of this process were painful, but, in general, it was an enjoyable experience.

I would also like to thank Dr. Michael McCauley for assisting in the editing process. It was also helpful to have someone that NPS actually considered "legitimate" while writing and routing the technical report.

While they must remain nameless, I would also like to thank the 11 brave individuals who either participated in the focus group or volunteered to be interviewed. The information gained from your assistance allowed me to test the NTSOD system using real-life data.

Finally, and most importantly, I would like to thank my beautiful wife, Alissa, and my lovely children, Kaidyn and Kilian. Gaining the experience required to write this thesis meant multiple deployments and extensive time away from home. Your love and support made those times much more bearable. Without you, this thesis would not be possible.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Human error is a major factor in accidents in the transportation and industrial fields. It is estimated to be a causal factor in 80% of mishaps (Reason, 1990). Surface warfare is no exception. Therefore, there is a need to develop valid methods for addressing human error in the Surface Warfare Officer (SWO) community. These methods will raise the overall awareness of SWOs. Once the community is aware of the effects of human error, safeguards can be established to limit the incidents.

B. OVERVIEW

> The Officer of the Deck (OOD) occupies a unique position in a naval ship. Nowhere in military or civilian life is there a parallel to the range and degree of responsibility that is placed in the hands of the OOD. As direct representative of the Captain, he or she acts with all the authority of command and, next to the Captain and the Executive Officer (XO), is the most important person on the ship. (Stavridis & Girrier, 2007, p. 1)

As demonstrated by the above quote from the *Watch Officer's Guide*, when underway, the OOD is the most important watch station onboard United States Naval surface vessels. OPNAVINST 3120.32C (2003) established the basic function of the watch. "The Officer of the Deck underway has been designated by the Commanding Officer (CO) to be in charge of the ship including its safe and proper operation (pp. 4–18)."

The current qualification process for OODs primarily consists of the completion of the Personnel Qualification Standards (PQS) and an oral board administered by the CO, XO, and selected department heads. The PQS for OODs consists of line items that qualified OODs verify once the candidate has displayed adequate knowledge of that item. These line items primarily concentrate on technical skills and knowledge such as visually determining a ship's target angle and determining compass error using a range. An example line item is, 'Set up a stadimeter and use it to determine the distance to a ship.' Once the candidate has sufficiently displayed the ability to use a stadimeter, a qualified

OOD will sign his or her PQS, thereby indicating that he or she is qualified in that specific area. The PQS also ensures that OODs have the technical knowledge required to respond to emergencies and infrequent tasks (e.g., mooring to a buoy, loss of steering). The PQS is the same across the Surface Fleet regardless of the type of ship. Therefore, it ensures baseline technical skills and knowledge for all OOD candidates.

While technical skills and knowledge are important in all high-risk organizations, including the military, Admiral Stavridis and Captain Girrier (2007) identified several other characteristics that are just as important for a successful OOD. Stavridis and Girrier state that forehandedness, vigilance, judgment, intuition, and leadership are traits that are just as valuable to OODs as technical knowledge and skills. These traits are examples of what psychologists have termed nontechnical skills.

Nontechnical skills are the "cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance" (Flin, O'Connor, & Crichton, 2008, p. 1). These skills are not mysterious or new to the surface warfare community. Nontechnical skills include skills such as leadership, communication, and decision making. The best OODs already possess these skills or develop them through experience. Nontechnical skills help watch standers respond to situations in an efficient and timely manner. They also help OODs handle situations that the current qualification process does not address.

Through observation, intuition, and experience, CO's develop an impression of the nontechnical skills possessed by individuals under their command. The CO, however, rarely has time to observe every watch stander for the time that is required to form an accurate assessment of the watch stander's competence. It is also difficult to translate intuition or a 'gut feeling' into meaningful data for making a decision as to whether the watch stander is competent. Unfortunately, there are no valid and reliable tools available to assess, and provide feedback on, the nontechnical skills of the OOD that are crucial for safe and effective performance.

C. PROBLEM STATEMENT

Technical advancements and increased redundancies in ships and other systems have reduced the chances of mechanical errors at sea that would be severe enough to cause major incidents. As a result, the majority of groundings, collisions, and other maritime accidents are a result of human error. A United States Coast Guard (USCG) report identified human error as a contributing factor to 75-96% of maritime casualties (Hetherington, Flin, & Mearns, 2006). Assessing and providing training and feedback on nontechnical skills could reduce these percentages. Providing COs with a mechanism for evaluating non-technical skills would allow them to identify potential shortfalls in the skills of OODs and train them more effectively. For example, a deficiency in situational awareness or decision making could be addressed and corrected during qualification before the individual ever stands a watch. Unfortunately, despite their recognized importance, commands do not consistently track or evaluate nontechnical skills for OOD candidates.

D. OBJECTIVES

There were two primary purposes to this thesis. The first was to identify the critical nontechnical skills for the OOD watch station. While Stavridis and Girrier (2007) recognized several important nontechnical skills, the traits they identified are difficult to observe and assess.

The second purpose was to develop a valid and reliable framework of non-technical skills for OODs so that they so that they can be observed and rated.

E. HUMAN SYSTEMS INTEGRATION (HSI) APPLICABILITY

HSI emphasizes human considerations as the main priority to reduce life cycle cost and optimize system performance. It is a multi-disciplinary field composed of eight basic domains: manpower, personnel, training, human factors engineering, system safety, health hazards, survivability, and habitability. This thesis focuses on addressing three of these domains as applied to the OOD watch station.

The first and primary domain addressed is training. Booher (2003) identified training as “the requisite knowledge, skills, and abilities needed by the available personnel to operate and maintain systems under operational conditions” (p. 3). This thesis identifies the nontechnical skills that every OOD should possess. The development of a valid and reliable system to evaluate the nontechnical skills of OODs will improve the training of OODs. The system will allow COs to identify areas of strength and weakness to assist potential OODs during their qualification process.

The second domain targeted with this thesis is personnel. Personnel is described as the “aptitudes, experiences, and other human characteristics necessary to achieve optimal system performance” (Booher, 2003, p. 3). This thesis identifies key nontechnical skills that can improve the performance of OODs. Since every individual, to varying degrees, already possesses these skills, providing a framework for measuring them could have impact on the type of person that selected to stand OOD during critical evolutions.

The last domain addressed is system safety. The above quote from Stravridis and Girrier demonstrates the immense amount of responsibility that an OOD has while on watch. An ineffective OOD possesses the ability to cause severe damage to the ship and equipment as well as injury to the ship’s crew. Improving the proficiency of the watch standers will increase the overall safety of the ships.

F. SCOPE AND LIMITATIONS

This thesis focused exclusively on the nontechnical skills critical for standing OOD on U.S. Navy surface warships. The method used to identify the skills is focused on critical or stressful moments that occur while standing the watch. However, the system applies to the full spectrum of OOD activities.

While some categories may fit, the findings should not be applied to similar watches on other platforms such as submarines or merchant ships without carrying out research to assess the validity of the system. The system as a whole also was not designed to address other watch stations such as Engineering Officer of the Watch (EOOW) or Tactical Watch Officer (TAO).

II. EVALUATING NONTECHNICAL SKILLS

A. MODELS OF HUMAN ERROR

James Reason (1990) estimated that human error causes 80% of accidents in high-risk organizations. He identified human error as “a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency” (p. 9).

Based upon research examining failures in high reliability organizations, as well as the work of other researcher (e.g., Rasmussen, 1974), Reason (1997) developed an organizational model of human error commonly described as the ‘Swiss Cheese’ model. This model allows for the identification of active failures and latent conditions in an organization. Active failures are unsafe acts that have a direct impact on the safety of the system. These acts are normally committed by the system operator (e.g., pilot, maintenance personnel, OOD). Latent conditions are the failures or oversights that allow or, in some cases, encourage the end user to commit the unsafe act that lead to the accident. Latent conditions include issues such as poor design, failures of supervision, and shortfalls in training. Latent conditions arise from decisions made by manufacturers, designers, and managers. They may be present in the system for many years before any mishaps occur (Reason, 1997). Although latent conditions are important contributors to accidents and poor performance, the focus of this thesis is on addressing the active failures of the Surface Navy, and more specifically the OOD.

Organizations whose performance may be catastrophically impacted by failures in complex human technology systems are known as high-risk industries. High Reliability Organizations (HROs) are the sub-set of high-risk organizations that succeed in avoiding catastrophes in high-risk environments (Roberts & Rousseau, 1989). The common factor underlying HROs is that, while a failure of reliability has the potential for death, loss, damage to assets, or ecological disaster, these organizations have developed unique properties that enable them to adapt to unexpected events. They develop this ability by

placing increasing importance on understanding and leveraging the role of the human operator (O’Connor & Cohn, 2010). Carter-Trahan (2009) identified U.S. Navy vessels as HROs. While the Navy has fewer accidents than would be expected due to the complex environment that they operate in, mishaps do still occur. This thesis increases the understanding of the human operators, in this case OOD, by identifying the nontechnical skills that are important to safe and effective operation.

Flin, O’Connor, and Crichton (2008) describe non-technical skills as “the cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance” (p. 1). Situational awareness, leadership, teamwork, and decision making are all examples of nontechnical skills. Over time, effective operators naturally develop these skills in order to function at high levels of performance. While human error cannot be eliminated, improvements in nontechnical skills can identify and mitigate potential risks.

B. BEHAVIORAL MARKER SYSTEMS

While identifying relevant and applicable nontechnical skills is crucial to safety, assessing the skills is also very important. Flin et al. (2008) define assessment as “the process of observing, recording, interpreting, and evaluating individual performance, usually against a standard defined by a professional body, a company, or a safety regulator” (p. 269). The primary technique for assessing or evaluating nontechnical skills is through the use behavioral marker systems.

Klampfer et al. (2001) defined behavioral markers as “observable, nontechnical behaviors that contribute to superior or substandard performance within a work environment” (p. 10). Behavioral markers derive from data analysis regarding performance that contributes to an outcome, either successful or unsuccessful. Behavioral markers could be used in any domain, but they are most frequently used in industries in which safety is a primary concern. Klampfer et al. (2001) identified five characteristics of a good behavioral marker.

1. It should describe a specific and observable behavior instead of an attitude or personality trait. The operator’s ability to delegate responsibilities is observable and a good indicator of leadership skills. However, the operator’s attitude towards leadership is not observable, so it would not make an effective behavioral marker.
2. While it does not have to be present in every situation, the behavioral marker should have a causal relationship with the performance outcome. In other words, the inability for the operator to communicate should result in poorer performance for his or her team.
3. It needs to use domain specific language. A behavioral marker that was developed for the maritime shipping industry is unlikely to be useful in a nuclear power plant control room due to maritime specific language.
4. It should employ simple phraseology. This attribute will allow the marker to be used and understood by a wider range of individuals including the assessor and the operator
5. It should describe a clear concept.

Organizations can use behavioral markers for multiple purposes. Behavioral markers can be used to give feedback on performance to individual operators and teams. They can highlight positive and negative examples of performance. The can also provide a common terminology for training and debriefing. The most common use for behavioral markers is as a tool to measure and evaluate performance (Klampfer et al., 2001).

Behavioral marker systems are rating tools that consist of multiple behavioral markers organized in a manner that allows an assessor to assign grades to each skill. The system normally consists of several broad category nontechnical skills (leadership, communications, etc.) that are then broken down into smaller elements or markers that are observable and relevant. The system should also contain a rating scale that allows for consistent assessment across different categories.

While Klampfer et al. (2001) identified the seven desirable characteristics for a behavioral marker, Flin et al. (2008) identified seven properties for an effective behavioral marker system as a whole: sensitivity, reliability, validity, structure, transparency, usability, and baselines.

- *Sensitivity* refers to assessor’s ability to distinguish between good and bad performance based upon the markers. If a behavioral marker system includes leadership as one of its nontechnical skills, it should be easy to identify what markers indicate good leadership and which indicate bad leadership.

- *Reliability* relates to the stability of the measurement. In other words, the same action should be given the same rating in multiple situations. There are three primary aspects to reliability: test-retest, internal reliability, and inter-rater reliability. Test-retest refers to the stability over time. Raters should make the same judgment on the same incident occurring at two different times. Internal reliability tests the level of agreement between individual markers that are intended to measure the same nontechnical skill. For example, three behavioral markers that are being used to measure communication skills should show correlated scores. Inter-rater reliability is the ability of different raters to give similar actions the same score (Flin et al., 2008).
- *Validity* is the extent to which a marker actually measures the nontechnical skill for which it was designed to assess. Behavioral markers should accurately reflect differences in performance. There are two basic aspects of validity: face validity and construct validity. Face validity is how appropriate the construct looks to the user. In other words, do the markers for decision making look like they are actually measuring decision making? Construct validity is whether the rating system is actually measuring what it claims. In other words, higher scores should actually relate to safer and more efficient performance (Flin et al., 2008).
- *Structure* refers to how well the components and markers are organized. The perfect behavioral marker system would address all possible behaviors and there would be no overlap between the categories. Unfortunately, it is impossible to create a system that can be used for every situation. There always will be some overlap between the categories (for example, situational awareness and decision making are intrinsically related). The goal, however, is to limit the overlap and increase the total coverage as much as possible (Flin et al., 2008).
- The system should also be *transparent* or understandable to the operators who are being rated. They should have information on the reliability and validity of the system, and they should know the criteria against which they are graded.
- The behavioral marker system needs to be *usable*. The framework should be simple and easy to understand. The targeted behaviors should be easy to observe, and it should not significantly increase the rater's workload. They system should also have domain-specific language that is familiar to both the assessor and operator.
- *Baselines* refer to how appropriate the performance criterion is for the experience level of the operators. More experienced users should be held to a higher standard than trainees or beginners (Flin et al., 2008).

Flin et al. (2008) also identified three significant limitations to behavioral marker systems. First, they are never capable of capturing every possible aspect of performance. The variables that effect performance are simply too wide ranging to rate with one system. The second limitation is that there may be limited opportunities to observe some behaviors. Important, but infrequent, capabilities such as conflict management or intuitive decision making may not occur often enough to rate on a consistent basis. The final limitation is the inherent limitations of human observers. While human raters bring experience and skills that a fully automatic assessor cannot match, they also bring their own biases and perceptions.

It is important to note that behavioral marker systems need to be domain-specific. O'Connor, O'Dea, and Melton, (2007) illustrated that the Human Factors Analysis and Classification System (HFACS), which was developed for military aviation, was not appropriate for assessing human error or nontechnical skills for U.S. Navy divers. O'Connor et al. then tried to apply a behavioral marker system designed for the offshore oil industry. Significant changes were made to the taxonomy before it could be effectively applied to U.S. Navy diving.

C. DOMAINS IN WHICH BEHAVIORAL MARKER SYSTEMS HAVE BEEN USED

Behavioral markers are expensive to develop and utilize given the level of training required for users. Consequently, they have mainly been developed for occupations where safety is prime and simulators are used for training and assessment (Flin et al, 2008). The following paragraphs will briefly discuss behavioral marker systems that have been developed for civil aviation, medicine, and research that has been carried out in the civilian maritime industry.

1. CIVIL AVIATION

a. University of Texas Behavioral Markers System

The first behavioral marker system for pilots was developed as part of the University of Texas Human Factors Research Project. The study had two primary

purposes. The first was to evaluate the effectiveness of crew resource management (CRM) by measuring observable behaviors. The second goal was to aid in the development of future CRM programs (Klampfer et al., 2001).

After the project produced the original set of behavioral markers, the University of Texas began collecting systematic data on all aspects of an airline’s operations. The markers were incorporated into a system known as the Line/LOS Checklist. As the use of the system grew, it became obvious that there were significant differences in crew behaviors during flight. The researchers then modified the system to address the markers for each phase of flight. The current system is shown in Figure 1 (Klampfer et al., 2001).

			Phase
SOP BRIEFING	The required briefing was interactive and operationally thorough	- Concise, not rushed, and met SOP requirements - Bottom lines were established	P-D
PLANS STATED	Operational plans and decisions were communicated and acknowledged	- Shared understanding about plans - "Everybody on the same page"	P-D
WORKLOAD ASSIGNMENT	Roles and responsibilities were defined for normal and non-normal situations	- Workload assignments were communicated and acknowledged	P-D
CONTINGENCY MANAGEMENT	Crew members developed effective strategies to manage threats to safety	- Threats and their consequences were anticipated - Used all available resources to manage threats	P-D
MONITOR / CROSSCHECK	Crew members actively monitored and cross-checked systems and other crew members	- Aircraft position, settings, and crew actions were verified	P-T-D
WORKLOAD MANAGEMENT	Operational tasks were prioritized and properly managed to handle primary flight duties	- Avoided task fixation - Did not allow work overload	P-T-D
VIGILANCE	Crew members remained alert of the environment and position of the aircraft	- Crew members maintained situational awareness	P-T-D
AUTOMATION MANAGENT	Automation was properly managed to balance situational and/or workload requirements	- Automation setup was briefed to other members - Effective recovery techniques from automation anomalies	P-T-D
EVALUATION OF PLANS	Existing plans were reviewed and modified when necessary	- Crew decisions and actions were openly analyzed to make sure the existing plan was the best plan	P-T
INQUIRY	Crew members asked questions to investigate and/or clarify current plans of action	- Crew members not afraid to express a lack of knowledge - "Nothing taken for granted" attitude	P-T
ASSERTIVNESS	Crew members stated critical information and/or solutions with appropriate persistence	- Crew members spoke up without hesitation	P-T
COMMUNICATION ENVIRONMENT	Environment for open communication was established and maintained	- Good cross talk – flow of information was fluid, clear, and direct	G
LEADERSHIP	Captain showed leadership and coordinated flight deck activities	- In command, decisive, and encouraged crew participation	G

1 = Poor	2 = Marginal	3 = Good	4 = Outstanding
Observed performance had safety implications	Observed performance was barely adequate	Observed performance was effective	Observed performance was truly noteworthy

Figure 1. University of Texas Behavioral Markers for Airline Pilots (From Klampfer et al., 2001)

b. Nontechnical Skills (NOTECHS) of Pilots

In 1996, legislation created a need for a generic method of evaluating the nontechnical skills of pilots that would be applicable across Europe (Kanki, Helmreich, & Anca, 2010). Any method that was developed had to be respectful to cultural and corporate differences and usable by airline instructors and examiners. A research team consisting of pilots and psychologists from around Europe developed the NOTECHS system to assess the CRM skills of individual pilots.

The researchers began by examining the behavioral marker systems already employed by a number of large European airlines. Each system had problems that did not allow the researchers to generalize it for all of European aviation. The systems were either too general, too specific, or they assessed the crew as a whole instead of individual pilots. As a result, the researchers developed a new taxonomy and rating system to assess pilot's nontechnical skills (Kanki, Helmreich, & Anca, 2010).

The researchers used an extensive literature review and an examination of existing behavioral marker systems. Airline captains with considerable experience in using behavioral marker systems then advised on the final design. The result is a NOTECHS system that consists of four nontechnical skill categories (cooperation, leadership and managerial skills, situational awareness, and decision making) with component elements or markers for each. Figure 2 shows the NOTECHS system (Klampfer et al., 2001).

Categories	Elements	Example Behaviours
COOPERATION	Team building and maintaining	- Establishes atmosphere for open communication and participation
	Considering others	- Takes Condition of other crew members into account
	Supporting others	- Helps other crew members in demanding situation
	Conflict solving	- Concentrates on what is right rather than who is right
LEADERSHIP & MANAGERIAL SKILLS	Use of authory and assertiveness	- Takes initiative to ensure involvement and task completion
	Maintaining Standards	- Intervenes if task completion deviates from standards
	Planning and co-ordinating	- Clearly states intentions and goals
	Workload management	- Allocates enough time to complete tasks
SITUATION AWARENESS	System awareness	- Monitors and reports changes in system states
	Environmental awareness	- Collect information about the environment
	Anticipation	- Identifies possible / future problems
DECISION MAKING	Problem definition / diagnosis	- Reviews casual factors with other crew members
	Option generation	- States alternative courses of action - Asks other crew member for options
	Risk assessment / Option choice	- Considers and shares risks of alternative courses of action
	Outcome review	- Checks outcome against plan

Very Poor	Poor	Acceptable	Good	Very Good
Observed behaviour directly endangers flight safety	Observed behaviour in other conditions could endanger flight safety	Observed behaviour does not endanger flight safety but needs improvement	Observed behaviour enhances flight safety	Observed behaviour optimally enhances flight safety and could serve as an example for other pilots

Figure 2. The NOTECHS Behavioral Marker System (From Klampfer et al., 2001)

2. MEDICINE

a. Nontechnical Skills (ANTS) of Anesthetists

Fletcher et al., (2004) developed a behavioral marker system for assessing the nontechnical skills of anesthesiologists. The performance of anesthesiologists was often recorded using video. However, there was no valid and reliable tool to rate the nontechnical performance of the anesthesiologists. Fletcher et al. developed the Anesthetists' Nontechnical Skills (ANTS) system to fill that void.

Fletcher et al. (2004) started by developing a prototype taxonomy using literature review and examinations of existing marker systems. They started with eight nontechnical skills: leadership, team working, communication, task management, situational assessment, situational awareness, decision making, and personal factors.

They then conducted 29 interviews using the critical incident technique (CIT) method. They were able to derive 116 statements about nontechnical issues from the transcripts. Fletcher et al. (2004) then used the interview statements, observations in theater, and anesthesia incident reports to trim down the list of nontechnical skills and individual elements or markers. The final categories were team management, team working, situational awareness, and decision making (see Figure 3).

Category	Elements
Task management	Planning and preparing
	Prioritising
	Providing and maintaining standards
	Identifying and utilising resources
Team working	Co-ordinating activities with team members
	Exchanging information
	Using authority and assertiveness
	Assessing capabilities
	Supporting others
Situation awareness	Gathering information
	Recognising and understanding
	Anticipating
Decision making	Identifying options
	Balancing risks and selecting options
	Re-evaluating

Figure 3. Nontechnical Skills and Associated Elements for the ANTS system (From Fletcher et al., 2004)

b. *Nontechnical Skills (NOTSS) of Surgeons*

Building on the work carried out by Fletcher et al. (2004), Yule, Flin, Paterson-Brown, Maran, and Rowley (2006) identified a training gap for surgeons. Analysis of adverse events in surgery revealed that many of the underlying causes were not technical failures. Behaviors such as communication were commonly found to be the

cause. While surgeons needed to demonstrate these nontechnical skills, training did not address them. Yule et al. (2006) created a behavioral marker system for surgeons entitled the Nontechnical Skills for Surgeons (NOTSS) system.

The researchers interviewed 27 surgeons utilizing the CIT method. They asked the surgeons to describe critical incidents in which they were involved. The authors used the interviews to identify the key nontechnical skills for surgeons. They were able to derive a list of 150 nontechnical skills from the interview transcripts. Yule et al., (2006) used the list to develop broad categories and associated elements. A group of consultant surgeons then reduced the structure into a taxonomy comprising five categories (situational awareness, decision making, communication, and leadership) and 12 elements. Figure 4 displays the NOTSS system (Flin, Yule, Paterson-Brown, Rowley, & Maran, 2006).

Hospital .... Trainer name .... Date ....

Trainee name .... Operation ....

Category	Category rating*	Element	Element rating*	Feedback on performance and debriefing notes
Situation Awareness		Gathering informatior		
		Understanding information		
		Projecting and anticipating future state		
Decision Making		Considering options		
		Selecting and communicating option		
		Implementing and reviewing decisions		
Communication and Teamwork		Exchanging informatior		
		Establishing a shared understanding		
		Co-ordinating team activities		
Leadership		Setting and maintaining standards		
		Supporting others		
		Coping with pressure		

* 1 Poor; 2 Marginal; 3 Acceptable; 4 Good; N/A Not Applicable

1 Poor	Performance endangered or potentially endangered patient safety, serious remediation is required
2 Marginal	Performance indicated cause for concern, considerable improvement is needed
3 Acceptable	Performance was of a satisfactory standard but could be improved
4 Good	Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others
N/A	Not Applicable

Figure 4. The NOTSS System (From Flin et al., 2006)

3. MARITIME SHIPPING

While no behavioral marker systems have been developed for the maritime industry, the Warsash Maritime Center in England has been conducting research into the feasibility of applying behavioral markers to the maritime industry (Gatfield, 2005). The research has three primary areas of focus. The first is understanding how behavioral markers can be used to assess the crisis management capabilities of engineering officers. The second is developing a behavioral marker assessment framework. The final focus is providing the maritime community with an understanding of how a behavioral marker system can be applied to the industry.

The research is being conducted by observing exercise scenarios within simulators. While the research is not yet complete, several behavioral markers have already been found that significantly affect the overall performance of engineers in the simulators. Figure 5 shows the identified behavioral markers and the associated nontechnical skills (Gatfield, 2005).

Behavioural Marker	Characterisation
Ratio of the degree of feedback control to the degree of predictive control.	Indication of the level of situational awareness.
The number of alternative hypotheses and actions communicated to team members.	An indication of teamwork and the building of a shared mental model.
Level of satisficing exhibited.	Considering only as many alternatives as needed to discover one that satisfies.
Communicating in a way that shares ones mental model.	Building, maintaining and refining the accuracy of the shared mental model of the team.
Relevance and timeliness of unsolicited information passed between team members.	A measure of the degree of congruence between the mental models held by individual team members.
Level of anticipation of other team members needs.	Indication of the level of situational awareness.
Level of anticipation of future actions and task requirements.	Indication of the level of situational awareness.
Focus is too much on the reduction of uncertainty.	Indication of a tendency towards analytical decision-making, and away from naturalistic decision-making.
Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system.	An indication of the lack of a situation overview.
Amount of sampling behaviour exhibited.	An indication of the updating of situational awareness and mental model.
Number of unfinished sentences.	A measure of uncertainty.
Delegation of work tasks.	A measure of the effective use of all team members, and the alleviation of overload.
Patterns of movement.	Interpretation of patterns of movement to determine degree of situation overview.

Figure 5. Behavioral Markers Identified for Merchant Marine Engineering Officers (From Gatfield, 2005)

D. CONCLUSION

Effective nontechnical skills are crucial for effective performance in high risk organizations. Behavioral markers have been shown to be a valid and reliable method for providing feedback to operators on nontechnical skills and are being used in a number of high-risk domains. Given the prevalence and success of behavioral marker systems, it is suggested that they may be an effective method for improving nontechnical skills in the

Surface Navy. The next chapter will provide evidence for why behavioral markers will benefit performance of Surface Navy personnel. The chapter also will explain why the OOD watch station is the first watch for which behavioral markers should be developed.

THIS PAGE INTENTIONALLY LEFT BLANK

III. THE NEED FOR A BEHAVIORAL MARKER SYSTEM FOR OODS

A. ACTIVE FAILURES IN THE MARITIME INDUSTRY

Dr. Anita Rothblum (2000) notes that the material conditions of the shipping industry have been constantly improving over the last 40 years. Improvements in hull design, ship stability, propulsion, and navigational aids all have been implemented to reduce casualties and increase efficiency. Unfortunately, despite these advancements, the maritime casualty rate is still high. Dr. Rothblum states that this result is due to maritime shipping being a people system, so human error is a large contributing factor. She estimates that human error causes 75–96% or maritime casualties. Studies have shown that human error contributes to 84–88% of tanker accidents, 79% of towing vessel groundings, and 89–96% of collisions (Rothblum, 2000).

Dr. Rothblum also reported the results of a Dutch study of 100 maritime casualties. The study found multiple causes for each accident. Of the 100 incidents, human error contributed to 96 of them, with 93 events having more than one human error. An important finding of the study was the fact that every human error was a necessary condition for the accident to occur. In other words, the accidents would not have happened if the human errors did not occur (Wagenaar & Groeneweg, 1987).

In order to illustrate human error in shipping, Dr. Rothblum (2000) breaks down two real-life maritime incidents. The first accident is the collision between the motor vessel (M/V) SANTA CRUZ II and the U.S. Coast Guard Cutter (USCGC) CUYAHOGA that resulted in the death of 11 Coast Guardsmen. The vessels were able to see each other visually, and they had radar returns. There were no mechanical malfunctions, and no severe environmental conditions. The sole cause of the accident was human error (Rothblum, 2000).

The captain of the CUYAHOGA made the first mistake. He misread the light configuration of the SANTA CRUZ II. Therefore, the captain did not have an accurate mental picture of the other vessel’s size or heading. As a result, the Captain ordered a

turn that took the CUYAHOGA directly across the bow of the other vessel. The second fateful error was the failure of the crew to notify the captain of his mistake. They realized what was happening, but they were either unwilling or afraid to question the captain's decision. Fatigue and excessive workload may have contributed to the errors committed by both the captain and crew because the cutter was undermanned (Rothblum, 2000).

The second incident that Dr. Rothblum examined was the grounding of the TORREY CANYON. This grounding took place in the English Channel in broad daylight and calm weather. It resulted in the spillage of 100,000 tons of oil. Four separate human errors contributed to the accident. The first two combined to put a large amount of pressure on the master of the vessel to make good time. The master believed that if he did not make his intended port by the next high tide, he may have to wait up to five days before pulling in. He also needed to transfer cargo in order to even the ship's draft. He did not want to perform the task underway because it would increase the chances of oil spilling on his decks and he did not want to have the appearance of a messy ship. Therefore, the master was in a hurry to pull into port.

The third human error was the master's decision to go through the Scilly Islands instead of around them in order to save time. Unfortunately, the captain was not familiar with the route, and he did not have the correct charts for the area.

The helmsman made the final mistake. He did not realize that the ship was on autopilot, so when he attempted to execute the turning order given by the master, nothing happened. By the time the mistake was discovered, it was too late to make the turn, and the vessel ran aground (Rothblum, 2000).

Dr. Rothblum reported the results of a study conducted by the U.S. Coast Guard on improving safety and performance through human factors principles. The study found that the three largest problems were fatigue, inadequate communication, and inadequate technical knowledge (Rothblum, 2000). The author believes that improvements to these areas could contribute to the prevention of future maritime casualties

David Gatfield (2005) used the story of the M/V GREEN LILY to help justify the need for a behavioral marker system to assess the competence of crisis management in the engine control room of merchant vessels. The GREEN LILY was underway in vicinity of the Shetland Islands in severe weather. At one point, a seawater supply line in the engine room broke and caused significant flooding. The engineers gained control of the flooding just as the engine stopped. The chief engineer incorrectly assumed that the flooding caused the engine to stop. Several unsuccessful attempts were made to restart the engine as the vessel slowly drifted towards the island of Bressay. The Shetland Coastguard dispatched three tugs, a lifeboat, and a helicopter to assist the vessel. In the end, the GREEN LILY ran aground and broke apart. The coastguard managed to rescue the entire crew except for one sailor who was swept away and lost at sea (Gatfield, 2005).

The investigation showed that the engine failure was actually due to the mechanical over speed trip. If the engineers would have simply reset the trip, the engine would have started and the disaster averted. There was also a lack of situational awareness and communication between the crew and the rescuers that further complicated the situation. While this situation began with a mechanical failure, human error compounded the situation and led to the actual grounding.

Hetherington, Flin, and Mearns (2006) noted that technical advances in navigational aids have reduced the occurrences of mechanical errors severe enough to cause significant incidents. However, these advances have revealed the extent to which human errors are responsible for collisions, groundings, and other maritime accidents. The authors identified three main levels of issues that lead to maritime accidents: design, personnel, and organizational. While all three are contributing factors, the personnel issues are considered the immediate causes of most incidents. Personnel issues include stress, situational awareness, and communication (see Figure 6 for a complete list). The authors did identify organizational and management issues as underlying causes to mishaps, similar to Reason’s Swiss Cheese model.

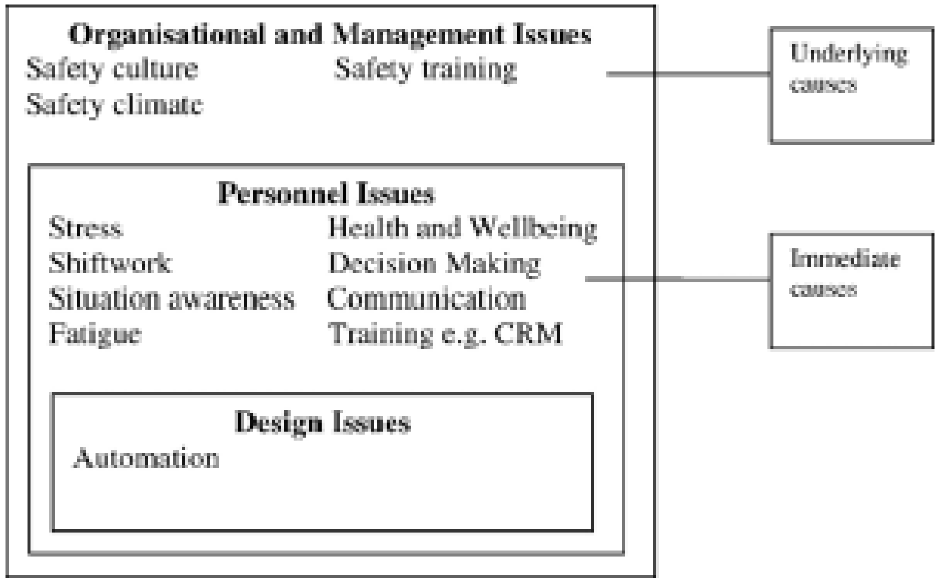

Figure 6. An Organizing Framework for Human Factors That Contribute to Accidents in Shipping (From Hetherington et al., 2006)

B. ACTIVE FAILURES IN THE U.S. NAVY

Although the U.S. Navy collects a large amount of detailed data on the human factors causes of aviation mishaps, the same is not true in the surface community. Carter-Trahan (2009) found that, in an examination of 111 major mishaps in the surface warfare community investigated by the Naval Safety Center from 1999 to 2009, only 23% of mishaps were attributed to human factors (the remaining were attributed to material causes 12%, and unknown causes 65%). A similar finding was reported in an examination 263 U.S. Navy diving mishap reports by O'Connor, O'Dea, and Melton (2007) where 70% of the mishaps were attributed to "unknown" causes, with only 23% attributed to human factors. These percentages are far below the 80% of mishaps that are attributed to human error in naval aviation and other high-risk environments. Therefore, it is suggested that the information collected on surface mishaps may be an underestimation of the extent to which human error contributes to afloat mishaps. The following paragraphs will review a few case studies of afloat mishaps in which human error was causal.

On May 17, 1987, the USS STARK was underway in the Arabian Gulf conducting operations to protect Kuwaiti and Saudi oil tankers. That evening, two Exocet missiles struck the ship. The first failed to detonate, but the second exploded and resulted in the death of 37 sailors. Miller and Shattuck (2004) examined the incident with respect to the Dynamic Model of Situated Cognition (DMSC).

The sensors onboard the STARK were limited due to adverse weather conditions, so they relied on the information relayed from an Air Force Airborne Warning and Control System (AWACS) aircraft. The Iraqi aircraft that fired the missiles was automatically identified as friendly because it was a French-made Mirage fighter. At one point, the Iraqi fighter made a sharp turn and increased speed towards the STARK. While the watch standers perceived this information, it was not comprehended. As a result, the threat was not recognized (Miller & Shattuck, 2004).

Seven minutes later, the Iraqi aircraft turned directly towards the ship. Unfortunately, the crew did not detect this information. By the time the aircraft was actually queried, one missile had already been fired, and the other quickly followed. No sensor systems tracked the firing, and the first person to realize a missile was enroute was the forward lookout. Unfortunately, his information did not reach the combat information center (CIC) quickly enough, and the missile impacting the hull was the first that most of the ship knew of the attack. Failures in situational awareness and unsafe acts by the crew ensured that the ship was not able to respond to the attack. As a result, 37 crewmembers lost their lives.

On July 3, 1988, the USS VINCENNES mistakenly shot down Iranian Flight 655. During the time of the incident, the combat information center was extremely busy. The ship was tracking two separate groups of Iranian small boats, and CIC was also tracking the VINCENNES' helicopter, which was airborne, and an Iranian military P-3 aircraft also in the area. The official investigation reports that the ship was also dealing with a fouled gun mount and was maneuvering extensively in an effort to keep the other gun unmasked (Miller & Shattuck, 2004).

Flight 655 was initially identified as "unidentified assumed hostile." Shortly after take-off, the Identification Designation Supervisor received information that led him to believe the contact was actually an F-14 fighter aircraft. Although one officer did mention that the contact could be a commercial aircraft, it was officially designated as an F-14.

When Flight 655 was about 12–15 nautical miles away from the VINCENNES, the Tactical Information Coordinator mistakenly reported that the contact had started to descend. These mistakes were all human errors that ultimately led to the downing of Flight 655. The mistakes were caused by increased stress, decreased situational awareness, and a preconceived notion of the environment. If one or all of these human errors had been avoided, the result of the encounter may have been entirely different (Miller & Shattuck, 2004).

In her thesis, Carter-Trahan (2009) conducted a case study of the USS DWIGHT D. EISENHOWER ramming the Spanish Bulk Carrier URDULIZ while the Spanish ship was at anchor in the Chesapeake Bay. Multiple human errors contributed to the collision. While the radar navigation team was able to obtain fixes on the other vessel, the visual navigation team was unable to obtain a fix on the URDULIZ. This lack of cohesion between the two teams created a sense of confusion, especially since the visual navigation team had primary responsibilities. However, the Navigator and OOD continued to make recommendations despite the disconnect in information. These recommendations included course and speed changes that were not communicated to the CO prior to or after execution. These mistakes, combined with a lack of communication from the harbor pilot and the watch standers onboard URDULIZ, caused the EISENHOWER to ram the other vessel.

While data from the Safety Center does not necessarily reflect a high amount of human error in Naval incidents, evidence from case studies suggest otherwise. Therefore, addressing the nontechnical skills of the SWO community is likely to be effective in improving safety and performance. The following section outlines why the initial focus should rest on the OOD watch station.

C. RATIONALE FOR THE FOCUS ON OOD

OPNAVINST 3120.32C establishes the basic function, duties, responsibilities, and authority of the OOD. The function of the OOD is simply stated: “The OOD underway has been designated by the Commanding Officer to be in charge of the ship including its safe and proper operations” (pp. 4–18). The instruction then lists the 18 duties, responsibilities, and authorities that the watch station demands. There is a wide variety of actions the OOD must perform. He or she must be aware of the tactical situation and geographic factors that may affect safe navigation. The OOD shall make all required reports to the CO. He or she must supervise the personnel on watch on the bridge and carry out the routine of the ship as published in the plan of the day. The OOD needs to supervise transmissions on all radio circuits and conduct on-the-job training for the other officers and enlisted personnel on the bridge. This list is just a small sample of the immense and varying responsibilities that are inherent to the OOD watch station (for the complete list of duties, please refer to OPNAVINST 3120.32C, pages 4–18 through 4–20).

The OOD reports directly to the CO for the general operation of the ship and to the Executive Officer (XO) for the ship’s routine. There are also certain situations where the OOD may report to the Navigator. OPNAVINST 3120.32C identifies 13 separate watch stations that report directly to the OOD. Figure 7 shows a simplified version of the underway watch structure.

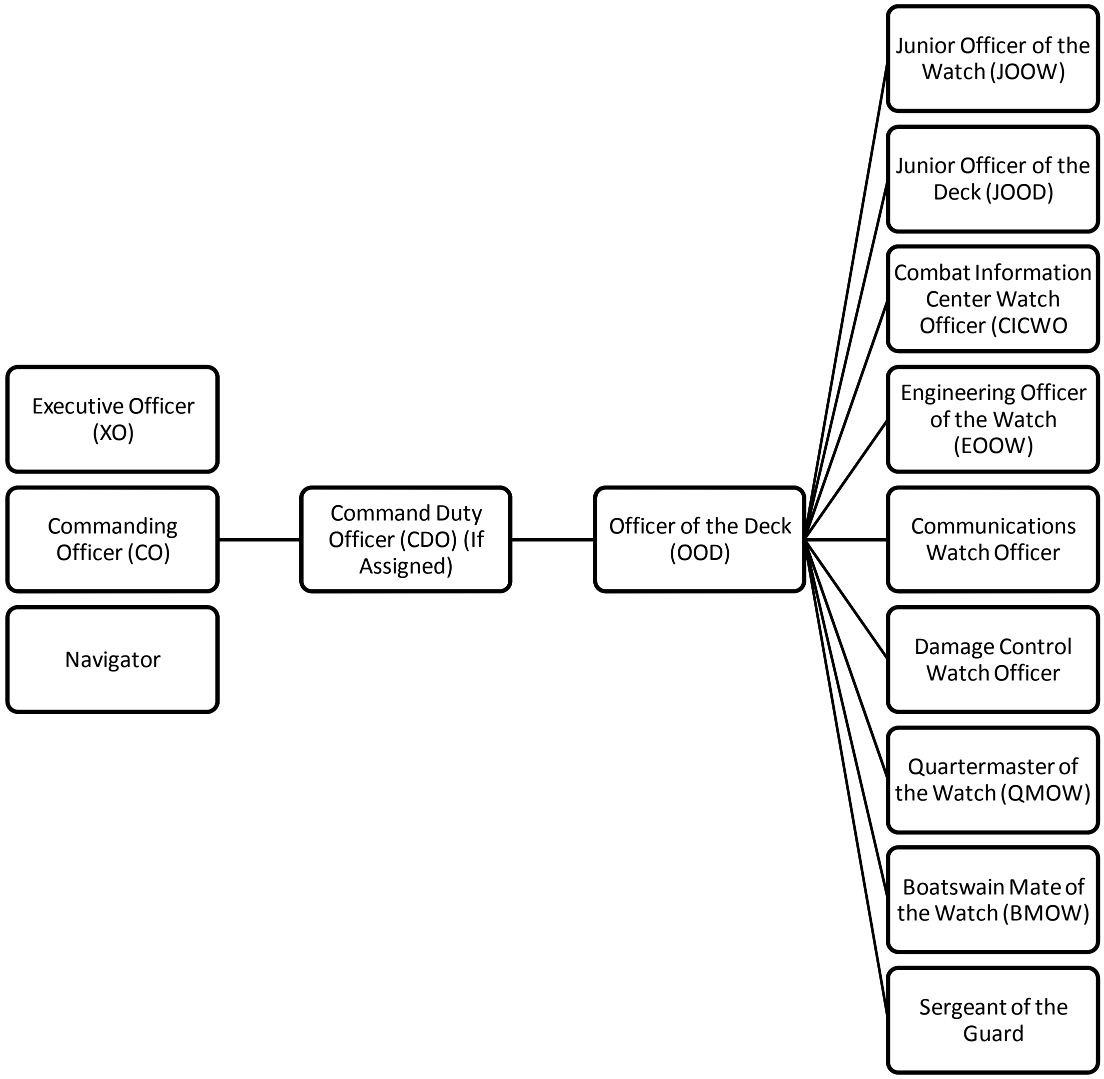

Figure 7. Underway Watch Structure
(Adapted from OPNAVINSTR 3120.32C page 4–38).

Admiral Stavridis and Captain Girrier (2007) discuss the importance of the OOD watch station. “Nowhere in military or civilian life is there a parallel to the range and degree of responsibility that is placed in the hands of the OOD (p. 1).” Stavridis and Girrier feel that the duties and responsibilities delineated in OPNAVINST 3120.32C are just the beginning. Special missions and circumstances can add more responsibilities, but they can never be reduced from what is stated in the instruction.

The OOD is accountable to the CO for every event that occurs during his or her watch. “As the captain’s direct representative, the OOD is the only person on board who can make decisions that affect the safety of the ship and the lives of her crew (Stavridis & Girrier, 2007, p. 3).” The CO cannot possibly be on the bridge at all times. In his absence the OOD works under his authority and speaks with his or her voice.

The OOD is the most important watch station onboard U.S. Navy surface ships. The responsibilities of the watch are wide ranging, and his or her decisions affect the entire crew. Therefore, observing and assessing the nontechnical skills of the OOD may improve the overall safety and effectiveness of the ship as a whole. While many different watch stations and positions could be analyzed, addressing the nontechnical skills of the OOD will go the furthest towards addressing human error in the Surface Navy.

D. CONCLUSION

Chapter II illustrated how a number of high-risk organizations have already recognized the need to address active failures and nontechnical skills of operators by using behavioral marker systems. This chapter focused on establishing the need to address failures of nontechnical skills on U.S. Navy surface ships. Unfortunately, it is not possible to design a system that will assess the nontechnical skills of an entire ship’s crew, nor can one thesis develop a framework of nontechnical skills for every watch station. Therefore, in the following chapter a study is carried out to develop a behavioral marker system for the OOD.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. METHOD

A. INTRODUCTION

The literature provides three primary methods for identifying nontechnical skills: event-based analyses, questioning techniques, and observational techniques (Flin et al., 2008). Event-based analyses consist of examining accident/near miss reports and/or examining information collected for confidential reporting systems. Accident and near-miss assessments entail analysis of past mishaps that have been recorded by the organization. The advantage of this method is that an actual incident has occurred that was severe enough to warrant investigation. Therefore, actual examples may be found where certain nontechnical skills were lacking and caused the accident. The disadvantage of accident analysis is that the reports may be incomplete and only represent part of the overall incident (Flin et al., 2008).

Confidential reporting systems collect information from workers about their mistakes or safety concerns that they would not normally be report. When this data are successfully fed back to leadership and management, safety improvements can be made. Unfortunately, individuals may be reluctant to report their own mistakes or concerns even when the system is confidential. However, organizations such as NASA and British Airways have successfully used confidential reporting systems to obtain information on unsafe events (O’Connor & O’Dea, 2007).

The second major system for identifying nontechnical skills is questioning techniques. This methodology can be divided into interviews, focus groups, and questionnaires. There are a three main interview types: structured, semi-structured, and unstructured. Interviews are advantageous because the interviewer is able to ask about the information that is relevant to him or her instead of hoping that an accident report contains pertinent data. The main disadvantage of interview techniques, however, is that the interviewee is only going to provide the information that he or she is comfortable giving and is able to verbalize. If the person is reluctant to incriminate themselves or others, the interview may not provide accurate or relevant information.

Focus groups involve a group of people providing feedback on a product or topic (Flin et al., 2008). The interaction between group members creates higher levels of discussion on specific topics. The advantage is that concentrated amounts of information can be gathered, especially on events that are not directly observable. The disadvantage is that the moderator or a powerful group member can control the discussion to the point that the data are no longer accurate.

Organizations can use questionnaires to collect specific information on nontechnical skills such as leadership, teamwork, and situational awareness. They have the advantage of being inexpensive and more efficient when collecting data from large groups of people. Unfortunately, there are limitations to the amount and type of information that can be collected.

The third technique that has been used to identify nontechnical skills are observations. Researchers can use observations to collect information about the workplace settings or a potential problem (Flin et al., 2008). Observations can be divided into three types: direct, participant, and remote. The advantage of observation is that the researcher is able to witness the information first-hand. Unfortunately, there is no guarantee that the desired action will happen, and it is possible that the researcher can alter the behaviors of the operators through his or her presence.

After considering these different methodologies, the author decided to use a combination of a literature review, a focus group, and interviews to identify the nontechnical skills of OODs. There are multiple reasons for this approach. First, as discussed in the previous chapter, the current mishap classification system does not adequately capture human error (Carter-Trahan, 2009) and no near-miss, or confidential reporting system exists. Since this thesis focused on the performance of OODs during stressful situations, observational methods also were not appropriate. Due to the rarity of such events in the surface navy, there is no guarantee that a researcher would have been able to observe the desired situation. However, the Naval Postgraduate School contains a wealth of knowledgeable and experienced OODs. The author took advantage of the presence of these experienced OODs to lead a focus group and interview individuals about situations they had encountered in the past while standing the watch as an OOD.

There is precedent for this method as the literature review provided multiple examples of researchers using subject matter experts to help create and refine their behavioral marker systems (Fletcher et al., 2004; Yule et al, 2007).

B. LITERATURE REVIEW

1. Purpose

The purpose of the literature review was to create an exhaustive list of categories of nontechnical skills that have been shown as necessary for effective performance in high-risk domains, with a particular focus on any research that had been carried out in the maritime industry. A comprehensive review provided a framework to begin production of the initial taxonomy of nontechnical skills. Textbooks, articles, and reports about other behavioral marker systems established a starting point for identifying the nontechnical skills that are applicable to the OOD watch station.

2. Method

The author consulted documentation to compile a list of nontechnical skills used in other fields. The only reference to behavioral markers in the maritime industry was by Gatfield (2005). He used simulator observations to develop a set of markers for assessing the competencies of merchant marine engineering officers during crisis management. However, the behavioral markers were not developed at the time of publication (nor are they discussed in any subsequent publications). Given the dearth of research on nontechnical skills in the maritime industry, four behavioral marker systems that were available in the literature (LLC, NOTECHS, ANTS, and NOTSS) were examined. Also included were the nontechnical skills identified by Stavridis and Girrier (2007), the skills discussed by Flin et al. (2008), and the relevant personnel issues identified by Hetherington et al (2006). This literature review resulted in the identification of 17 categories of nontechnical skills.

3. Results

The literature review created the 17 nontechnical skills shown in Table 1.

• Situational awareness	• Judgment
• Decision making	• Intuition/experience
• Communication	• Energy
• Team working	• Co-operation
• Leadership	• Management skills
• Managing stress	• Task management
• Coping with fatigue	• Workload management
• Forehandedness	• Assertiveness
• Vigilance	

Table 1. Initial List of Nontechnical Skills

4. Discussion

The author used the literature review to lay a foundation of nontechnical skills that could be considered applicable to the OOD watch station. The list included many skills that were overlapping or redundant, but it provided an excellent starting point for identifying which of the skills were important to the OOD.

C. FOCUS GROUP

1. Purpose

Several of the nontechnical skills were highly correlated, redundant, or not applicable to the OOD watch station. Therefore, the challenge was to identify which of the nontechnical skills in Table 1 were appropriate for assessing OODs. Once this had been carried out, the goal was for the SMEs to then agree upon an initial taxonomy of skills sub-divided into specific elements (the same structure as NOTECHS).

2. Method

a. Participants

The focus group consisted of four U.S. Navy junior officers led by an experienced OOD (the author). Each participant was a Surface Warfare Officer (SWO) attending the Naval Postgraduate School. All of the participants had been qualified as an

OOD for a minimum of a year. They had all stood the watch numerous times and spent time training others to do the same. They were also Human Systems Integration Masters' students, and had attended classes on human factors, safety, individual differences, and team working.

b. Procedure

The author presented each participant with the comprehensive list of nontechnical skills, definitions of each skill, and examples of behavioral marker systems. The participants were asked to familiarize themselves with the material prior to meeting. The author explained to the participants that the goal was to trim the initial list of skills to make it applicable to the OOD watch station. The secondary goal was to identify individual elements for each skill. The author encouraged open discussion among the participants and answered any questions that arose.

The group began by eliminating nontechnical skills that they felt were either redundant or not applicable. For example, the participants removed judgment because they felt it was included under the decision making category. Co-operation was eliminated because the focus group felt that is was not applicable since the study was addressing individual instead of team skills. The participants removed other skills such as forehandedness because the meanings were ambiguous and difficult to rate.

Once the focus group had identified a list of nontechnical skills, they were asked to use these to formulate a taxonomy of categories and associated elements. The participants were encouraged to draw from their experience as an OOD to identify specific behavioral markers skills. Following the discussion, the group reached a consensus on five categories, each with two or three corresponding elements. The taxonomy that resulted from the focus group is shown in Table 2.

3. Results

The focus group trimmed the list of 17 nontechnical skills down to five: leadership, decision making, situational awareness, communications, and stress management. The participants also identified two or three elements for each skill.

Because this is only the initial taxonomy and alterations are expected, definitions and descriptions of the categories will not be provided here. Please see the results section following the inter-rater reliability testing for descriptions of the elements in the final taxonomy. See Table 2 for the complete initial taxonomy of skills and elements.

Nontechnical Skill Category	Elements
Leadership	Establishing Authority
	Managing Workload
	Maintaining the Standards of the Watch
Decision Making	Defining the Problem
	Generating Possible Solutions
	Implementing Best Solution
Situational Awareness	Actively Gathering Information
	Responding to Changes in Information
	Anticipating Future Events
Communication	Selecting Correct Medium
	Sending Information Clearly and Concisely
	Effectively Receiving Information
Managing Stress	Maintaining Concentration
	Coping with Stressors

Table 2. Initial OOD Nontechnical Skills Taxonomy

4. Discussion

The author took advantage of the abundance of experience and knowledge located at the Naval Postgraduate School to assist in the identification of the applicable nontechnical skills. The focus group took the initial list of skills established from the literature review and specifically refined it for the OOD. The participants also identified the individual elements that apply to each skill category. This method produced the initial OOD nontechnical skills taxonomy shown in Table 2.

D. CRITICAL INCIDENT INTERVIEWS

1. Purpose

While the initial taxonomy appeared to be relevant and applicable to the OOD watch station, the author decided to test its validity using real-life examples. The author

interviewed qualified OODs regarding their experiences while standing the watch during stressful situations. These interviews provided the author with data from real situations. This data was used to conduct inter-rater reliability testing to identify potential issues with the taxonomy. Section E of this chapter discusses this testing.

2. Method

a. Participants

Eight individuals volunteered to participate in the interview process. Each person was a Surface Warfare Officer attending Naval Postgraduate School. Each participant had qualified as OOD at a previous command and spent significant time standing the watch and training others. Seven of the volunteers had recently completed their division officer tours while the remaining individual was on the shore duty following his department head tours.

b. Procedure

The author interviewed each volunteer separately. Each participant was asked to recount a stressful or otherwise memorable series of events that took place while he or she was standing OOD. The author encouraged each participant to describe as many experiences as they were comfortable providing. The interviews lasted between 25 and 60 minutes with an average of 45 minutes. The author utilized the Critical Incident Technique (CIT) to conduct the interviews.

The CIT was first developed and used by Flanagan (1954) to aid in pilot selection. However, the CIT technique has been widely used in studies of human error and safety (Kirwan & Ainsworth, 1992). The CIT enables the researcher to identify the (often tacit) knowledge of skills and expertise possessed by respondents by asking them to describe a challenging incident. It goes beyond procedural knowledge by probing the behavioral aspects of experience (O’Connor et al, 2008). The CIT interviews were conducted in four sweeps, defined below, by a qualified OOD knowledgeable about human factors. Each sweep was designed to jog the participant’s memory and get as much applicable and accurate information as possible.

Sweep 1 - Prompting the interviewee to identify a relevant incident: Each participant was asked to select and describe an event that occurred when they were standing watch as the OOD which they found to be particularly challenging. They were asked to describe the event from their own perspective and to describe it in detail, stage by stage, as it developed.

Sweep 2 - Filling in gaps in the incident: The interviewer repeated the reported incident back to the respondent, in order to check understanding. The respondent was told they should correct any mistakes in the account or add any information that was omitted during the recounting. This sweep helps to pinpoint gaps, both in time and events, and typically aids in recall of the missing portions.

Sweep 3 - Expanding on the incident to look for cues and factors affecting teamwork: The interviewer went through the event again, this time probing at various points and asking for more detailed description of the nontechnical aspects of the situation. This sweep involved questioning the reasoning process and looking for cues and rationale for the actions taken by team members.

Sweep 4 - “What if” queries: The interviewer asks questions about the participant’s perceptions, thoughts, judgments, and actions, and what would have happened if aspects of the scenario had been different. Each question is designed to extract more information that is applicable and gain a better understanding of the story as a whole.

c. Interview Transcription

The author took extensive notes in addition to obtaining an audio recording of each interview. Rather than transcribing the whole interview, the author used the notes and recordings to create a single, full report of the experience. The events were put into chronological order and repetitions were omitted. The author then edited the transcripts into a single format that was concise and comparable across the separate interviews. The author then pulled significant events or “statements” from each transcript. These statements were primarily actions that the OOD took while standing the watch. Each statement would be observable and ratable.

3. Results

Each of the eight participants related between one and three situations from their experience for a total of 16 scenarios. After transcribing the interviews, the author was able to extract 149 separate and distinct statements from the transcripts concerned with the nontechnical skills of the OOD. Exemplar statements are listed below.

- "The OOD went back to the bridge wing to look at the contact again."
- "The OOD reported both contacts to the CO prior to being relieved."
- "Once the ship was officially called into its waiting station, the OOD gave the order to proceed."
- "The OOD never called the oiler to state his intentions."
- "Once the ship was only 500 yards off the oiler's bow, the OOD ordered the Conning Officer to match the oiler's course and speed to avoid a collision."
- "The OOD did not use the wind envelope manual to prepare for helicopter operations."
- "The OOD continued to look at the radar scope instead of looking out the windows to see if they were approaching the oiler too fast."
- "The OOD did not know that the ship was four miles left of track."
- "The OOD ordered the bridge team to stop approaching the other vessel."
- "The OOD told the XO that they were not approaching the correct vessel."

4. Discussion

The scenarios provided ample information to evaluate the validity of the nontechnical skills taxonomy. The data reflected real-life situations experienced by qualified OODs during stressful or trying times. The next section discusses how the statements were used to develop a valid and reliable nontechnical skills taxonomy.

E. INTER-RATER RELIABILITY TESTING

1. Purpose

The author tested the inter-rater reliability to ensure that it was possible for two raters to reliably use the system to classify the nontechnical skills of OODs.

2. Method

Two individuals worked together to test the inter-rater reliability of the skills taxonomy. The first individual was the author who is a Surface Warfare Officer and qualified OOD. The second participant was a behavioral psychologist with experience in behavioral markers. He has produced similar products for other fields including nuclear power plant control rooms and U.S. Navy diving.

The inter-rater reliability testing consisted of classifying the statements extracted from the interviews by their applicable element in the skills taxonomy. The overall goal was for both raters to identify each statement by a single element and nontechnical skill. If each statement could be sorted into a category cleanly, the taxonomy was effective and no revisions were needed. However, if certain statements did not fit anywhere or applied to multiple categories, the system needed to be revised. The author calculated the effectiveness of the inter-rater reliability using Cohen’s kappa.

Cohen’s kappa (Cohen, 1960) is a method developed to calculate the inter-rater reliability between two individuals. It is considered a better measurement than the simple percentage of agreement because it takes chance agreement into account. Cohen’s kappa (κ) produces a value ≤ 1.0. The closer that the value is to 1.0, the better that the inter-rater reliability is said to be. Landis and Koch (1977) proposed that a kappa of less than 0 was indicative of poor agreement, between 0.0 and 0.20 indicates a slight agreement; between 0.21 and 0.40 a fair agreement; between 0.41 and 0.60 a moderate agreement; between 0.61 and 0.80 a substantial agreement; and between 0.81 and 1.00 almost perfect agreement. The goal of this study was to produce a $\kappa \geq .81$ between the raters. As will be shown below, four separate iterations of ratings were required before an acceptable level of kappa could be achieved.

a. Iteration 1

In the first iteration, the two raters classified 25 randomly selected statements using the initial taxonomy (see Table 2). The ratings were conducted independently. The first iteration resulted in an inter-rater reliability of $\kappa = .43$ indicating

moderate agreement. While this was a surprisingly good result for the first run, it was clear that changes needed to be made to the taxonomy to improve the inter-rater reliability.

Every nontechnical skill category showed some degree of disagreement, especially the individual elements of the leadership and managing stress categories. There was also confusion between the decision making, situational awareness, and communication skill categories. After discussion of areas of agreement and disagreement between the raters, the author made substantial changes to the taxonomy with the aim of improving it. Table 3 shows the revised taxonomy.

Nontechnical Skill Category	Elements
Leadership	Managing Workload
	Managing Stress
Decision Making	Considering Options
	Balancing Risks
	Implementing and Reviewing Decisions
Situational Awareness	Actively Gathering Information
	Responding to Changes in Information
	Anticipating Future Events
Communication	Issuing Orders
	Following Orders and Procedures
	Providing Information

Table 3. OOD Nontechnical Skills Taxonomy Revision 1

The author eliminated the "managing stress" nontechnical skill category and added it as an element under leadership. He also removed the two unclear elements from leadership. In an attempt to achieve some clarity between the other categories, the author created new elements for decision making and communication while keeping situational awareness the same.

b. Iteration 2

After revising the taxonomy, the raters tested it again. They classified 25 different statements using the revised taxonomy (see Table 3), and once again, the ratings

were done independently. The second iteration resulted in an inter-rater reliability of $\kappa = .65$ indicating substantial agreement. The improvement was expected, but it was apparent that portions of the system were still unclear and difficult to use.

The leadership category showed perfect concurrence, and there was good agreement on the communication elements. Unfortunately, there was still significant disagreement between decision making and situational awareness. The author addressed this problem by changing the elements of decision making because they were the least clear. Table 4 shows the second revision of the taxonomy.

Nontechnical Skill Category	Elements
Leadership	Managing Workload
	Managing Stress
Decision Making	Analytical Decision Making
	Following Orders and Procedures
	Intuitive Decision Making
Situational Awareness	Actively Gathering Information
	Responding to Changes in Information
	Anticipating Future Events
Communication	Issuing Orders
	Providing Information
	Receiving Information

Table 4. OOD Nontechnical Skills Taxonomy Revision 2

The author decided that broader elements would be easier to use and rate in the decision making category. It was decided to use the three naturalistic decision making styles (analytical, rule-based, and intuitive; Klein, 2008). However, the wording of the rule-based decision making was revised to better fit the OOD environment by calling it “following orders and procedures.” Because this element had previously been classified under communication, the author also had to make minor changes to that category.

c. Iteration 3

The raters then took 25 new interview statements and tested the reliability of the second revision of the taxonomy shown in Table 4. The third iteration resulted in an inter-rater reliability of $\kappa = .70$ indicating substantial agreement. While there were some minor improvements, a few problems still existed in the taxonomy.

The leadership and decision making categories showed very good agreement. However, there was some confusion between communication and situational awareness. Specifically, it was difficult to decide whether the OOD was gathering information or receiving information. In the example of a lookout making a report, the OOD was certainly receiving information, but he was also improving his situational awareness. The author decided that while both elements could apply to the situation, gathering information in the situational awareness category covers a wider range of events and is more useful. Therefore, the “receiving information” element was eliminated from the communication category. The situational awareness elements also were reworded to provide further clarification. Table 5 shows the final skills taxonomy.

d. Iteration 4

Due to the relatively minor changes made to the taxonomy in the third revision, the author decided to rate all 149 interview statements to properly test the latest version of the taxonomy. The fourth iteration resulted in an inter-rater reliability of $\kappa = .91$ indicating nearly perfect agreement. As the level of inter-rater reliability was higher than 0.81 (perfect agreement as defined by Landis & Koch, 1977), no further changes were made to the taxonomy.

3. Results

Table 5 depicts the final taxonomy of nontechnical skills and the distribution of the statements across the skills and elements. The inter-rater reliability testing created a system with four nontechnical skill categories: leadership, decision making, communications, and situational awareness. The following sections discuss each category along with the associated elements.

Skill	Element	Definition	Example	Skill %	Element % within skill	Overall element %
Leadership	Managing Watch Team	Effectively setting and maintaining the standards of the watch team.	*The OOD utilized the dead time in the schedule to review the emergency procedures with the helmsman.*	12.1	44.4	5.4
	Coping with Stress	Retaining a calm demeanor when under pressure and demonstrating to the watch that one is under control.	*Despite the added pressure from the XO, the OOD managed the stress and performed proficiently.*		55.6	6.7
Comms	Providing Information	Passing information along to other watch stations throughout the ship, as well as other assets in the area.	*The OOD called the other ships in formation to inform them that the passing oiler was dimly lit and difficult to see.*	28.2	33.3	9.4
	Issuing Orders	Effectively giving orders to other members of the watch team and other individuals as required.	*The OOD ordered the Engineering Officer of the Watch to start another engine.*		66.7	18.8
Situational Awareness	Gathering Awareness	Actively gathering information to keep up with the changing situation.	*At two nautical miles out, the OOD visually inspected the contact through binoculars.*	39.6	66.1	26.2
	Understanding Awareness	Achieving an understanding of what the available information means.	*The OOD identified the contact as a fishing vessel by analyzing the lighting configuration.*		25.4	10.1
	Anticipating Future Events	Forward planning in order to anticipate possible future problems.	*The OOD had the Conning Officer drive slightly right of the intended track because he knew the wind and current would push the ship to the left.*		8.5	3.4
Decision Making	Analytical Decision Making	Generating and comparing multiple courses of actions to come up with the optimal solution.	*Once the oiler was located, the OOD decided to start driving towards her early to ensure that there was plenty of time to set up later.*	20.1	53.3	10.7
	Following Orders & Procedures	Following documented procedures or direct orders from superior officers.	*The OOD used the wind envelope guide to make sure that the winds were sufficient to conduct flight operations.*		40.0	8.1
	Intuitive Decision Making	Making quick decisions based upon prior experience and intuition.	*When the mysterious light finally materialized as a sailboat 300 yards off the port bow, the OOD immediately ordered "Hard Right Rudder."*		6.7	1.3

Table 5. Nontechnical Skills and Elements Percentage Table

a. Leadership

The Navy defines leadership as the "sum of those qualities of intellect, of human understanding, and of moral character that enable a person to inspire and to manage a group of people successfully" (Stavridis & Girrier, 2007, p. 8). The inter-rater reliability testing modified the original three elements shown in Table 2 down to the two shown in Table 5: managing watch team and managing stress.

A normal underway watch can be characterized by long stretches of monotony followed by periods of fast-paced action. Managing the workload refers to the OOD's ability keep the watch team functioning through the peaks and valleys of activity.

Life at sea can be a stressful environment for Surface Warfare Officers. The responsibilities to his or her division and collateral duties pile up, and fatigue often becomes an issue. The ability to block out, or compartmentalize, the stressors and concentrate on the watch is crucial to the effective performance of the watch team. For this reason, managing stress is a crucial element in the leadership of the OOD.

b. Decision Making

Decision making can be defined as "the process of reaching a judgment or choosing an option to meet the needs of a given situation" (Flin et al., 2008, p. 41). While the focus group developed three broad elements of decision making, the elements did not allow for acceptable levels of inter-rater reliability. As a result, three new elements were created based on Klein (2008): analytical decision making, following orders and procedures, and intuitive decision making.

Analytical decision making involves the comparisons of multiple courses of actions to develop the optimal solution. The OOD encounters many situations that require a careful analysis of multiple options. The ability to consistently select the best alternative is a valuable skill for an OOD.

The OOD deals with many situations that are highly structured through either documented procedures or direct orders from superior officers. When properly

utilized, night orders, checklists, and emergency operating procedures are a few examples of the tools that make the OOD's job much easier. The OODs desire and ability to take advantage of these tools is an important element of decision making.

Intuitive decision making refers to situations where there is insufficient time to follow a documented procedure or conduct an analysis of alternatives. These instances require the OOD to make quick decisions based upon prior experience and intuition. These situations tend to be the most critical and the CO must have confidence that the OOD will be able to make the right decision.

c. Situational Awareness

Situational awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p. 36). As discussed earlier, the focus group created three elements for situational awareness based upon the levels of Endsley's (1995) model of situation awareness. These three elements survived the inter-rater reliability testing with only minor changes to the wording. The new elements are gathering awareness, understanding awareness, and anticipating future events.

The environment around a naval vessel is constantly in flux. The OOD must actively gather awareness about his or her surroundings to keep up with the changes and adjust accordingly. The OOD has many tools available to help increase situational awareness including radars, lookouts, and his or her own eyes. Properly utilizing these sensors and tools is crucial to gathering accurate awareness about the surrounding environment.

Simply gathering information is not enough to achieve good situational awareness. The OOD needs to use the tools at his or her disposal to achieve an understanding of what the information means.

Once there is a general understanding of the situation, effective OODs will be able to project the information into the future in order to anticipate what is coming next. Anticipating future events is crucial to keeping the ship out of dangerous environments and situations.

d. Communication

Communication is “the exchange of information, feedback or response, ideas, and feelings” (Flin et al., 2008, p. 69). The inter-rater reliability testing proved that the original elements of communication created by the focus group were largely ineffective. The new elements were: providing information and issuing orders.

The OOD is directly responsible for providing information to the CO and XO as well as the navigator in certain situations. He or she must pass information along to other assets in the area, as well as to other watch stations throughout the ship. The ability to provide the information in a clear and concise manner greatly increases the effectiveness of the OOD and the confidence of the chain of command.

The OOD is directly in charge of a watch team consisting of at least eight sailors. The engineering and combat watches also report to the OOD during normal operations. These responsibilities necessitate that the OOD issues orders to a wide range of individuals in many different situations. Therefore, issuing orders is an important skill for the OOD.

4. Discussion

Examining the NTSOD skills and elements that were most commonly used to categorize the interview statements provides evidence as to the nontechnical skills that should be emphasized during OOD training. From Table 5, it can be seen that the most frequently used elements were ‘issuing orders’ and ‘gathering information’. These two elements were used to categorize almost half of the interview statements.

Assertiveness and giving direct and explicit orders would appear to be an important nontechnical skill that should be emphasized in training junior OODs. This finding is not unexpected given that the OOD is the leader of the bridge team and must provide direction to the 13 different watch stations that report to him or her.

Gathering information was the most commonly used NTSOD element in this study. It was used to classify 26.2% of the interview statements. The failure of U.S. Navy sailors to gather information was a key causal factor in the ramming of the Spanish freighter by the *USS Eisenhower* (National Transportation Safety Board, 1990). In Endsley's (1995) three stage model of situational awareness, the first stage of gathering information is where most errors occur. Jones and Endsley (1996) examined situational awareness errors in civil aviation, while Sneddon, Mearns, and Flin (2006) examined them in offshore oil drilling. It was found that the most situational awareness errors occurred while gathering information (76.3% and 67.0% respectively), as opposed to understanding the meaning of the information (20.3% and 20.0% of errors respectively), or anticipating future states (3.4% and 13.0% of errors respectively). Therefore, given the importance to the OOD of the skill of gathering information, in combination with it being where most errors occur, the OOD must be given training that emphasizes and provides practice in effective information gathering (e.g., maintaining a good scan; see Flin et al, 2008, for more details).

Intuitive decision making was rarely used to classify the interview statements (it was only used to classify 1.3% of the statements). Part of the explanation for this surprising finding may be that the operations of a U.S. Navy ship are highly proceduralized. However, although it may be rare, the ability to make intuitive decisions is crucial for the OOD in risky, fast moving, or non-normal situations for which there is no procedure.

F. PROTOTYPE BEHAVIORAL MARKER SYSTEM

1. Purpose

The final stage in the development of the Nontechnical Skills for Officers of the Deck (NTSOD) rating system was to use the categories and elements shown in Table 5 to develop a nontechnical skills rating form that could be used by the fleet.

2. Method

In order to turn the nontechnical skills taxonomy into a rating form, the author studied multiple behavioral marker systems. Each had its strengths, but the format of the NOTSS system (Figure 6) showed the most potential for use in the Surface Navy environment and had received acceptable usability ratings from surgeons (Yule et al, 2006). The NOTSS system employs a simple format similar to that used in other behavioral marker systems. The header section allows the assessor to personalize each form to the individual being assessed and the particular evolution that they are performing. Another useful feature of the NOTSS system is the amount of space available for taking notes. The notes enabled by this space allow the assessor to provide a higher level of feedback to the trainee.

The NOTSS behavioral marker system was developed for surgeons, so it did need to be adapted before it could be used to assess OODs. The author changed the information in the header section to apply to standing watch. He included areas to write the ship name, trainee and assessor name, and the date. The author also felt that it was important to include spaces for the watch (2200–0200, 0200–0700, etc.) and evolutions (underway replenishment, normal underway steaming, etc.).

The author chose to use a four-point (unsatisfactory, marginal, satisfactory, outstanding) rating scale. The four-point scale allows for a higher degree of discrimination than a two- or three-point scale. The even number also forces the assessor to make a judgment about the trainee’s performance. There will be times when not every skill is represented during a watch. Therefore, the author did include the choice of not observed (N/O).

The author created a body similar to the NOTSS system that allows the assessor to give scores for both the individual elements and the nontechnical skills as a whole. This structure allows the trainee to get feedback regarding the individual skills that need improvement, but it also allows the CO or XO to get a quick idea of the performance by just looking at the broader categories. The author called the completed behavioral marker system the Nontechnical Skills for Officers of the Deck (NTSOD) rating form. Figure 8 depicts the proposed NTSOD rating form. Detailed instructions on the use of the NTSOD system are not provided here, but can be found in Long, O'Connor, & McCauley (2010).

3. Results

Figure 8 shows the completed NTSOD Rating Form.

Ship______________________	Trainee ____________________	Watch ____________________
Date______________________	Assessor___________________	Evolution __________________

Category	Category Rating*	Element	Element Rating*	Notes
Leadership		Managing Watch Team		
		Coping with Stress		
Communications		Providing Information		
		Issuing Orders		
Situational Awareness		Gathering Awareness		
		Understanding Awareness		
		Anticipating Future Events		
Decision Making		Analytical Decision Making		
		Following Orders & Procedures		
		Intuitive Decision Making		

* 1 – Unsatisfactory; 2 – Marginal; 3 – Satisfactory; 4 – Outstanding; N/O – Not Observed

1 – Unsatisfactory: Watchstander could endanger ship and crew without considerable improvement.
2 – Marginal: Watchstander requires improvement.
3 – Satisfactory: Watchstander performed at an acceptable level, but room for improvement exists.
4 – Outstanding: Watchstander performed at a consistently high level.
N/O – Not Observed: Element or skill was not observed during this evolution.

Figure 8. Nontechnical Skills for Officers of the Deck (NTSOD) Rating Form

4. Discussion

The prototype Nontechnical Skills for Officers of the Deck (NTSOD) taxonomy, shown in Figure 8, provides a format for evaluating the nontechnical skills of OODs that was based upon a sound research foundation. However, further developmental work is required to ensure that the system can be reliably used by evaluators. Once a prototype behavioral marker system has been developed, the next stage is for representative end users to rate standardized video enactments of scenarios similar to those encountered in the actual environment (Fletcher et al., 2004; Yule et al., 2008). The purpose of this evaluation is to assess the sensitivity, inter-rater reliability, and internal reliability of the system. The scenarios are typically filmed in a simulator with subject matter experts acting the main roles.

Should an acceptable level of reliability and sensitivity be achieved, the usability of the behavioral marker system is then evaluated by evaluating the behaviors of trainees in the actual environment (i.e., flight deck, operating theatre, ship's bridge, etc.). Completed rating forms are examined to assess whether all the elements and categories are being used and information is obtained from the raters on the usability of the behavioral marker system.

The NTSOD taxonomy not only has implications for the evaluation of the nontechnical skills of OODs, but could also be used to provide a research foundation for the content of the U.S. Navy's bridge resource management (BRM) program. As stated by Hetherington et al. (2004) "a review of the literature reveals that there appears to be no empirical foundation for this type of course beyond research that was originally conducted in the formation of aviation [crew resource management]_CRM courses" (p. 407). The U.S. Navy's BRM program is no exception. Carter-Trahan (2009) carried out a study of the attitudes and knowledge of OODs regarding human factors issues that have been identified as causal to mishaps in high-risk organization. She stated that there was a need for a systematic research effort to identify the particular human factors issues that should be included as part of the Navy's BRM program.

The first stage in the development of any training program should be a needs assessment. In the context of nontechnical skills training, such as BRM/CRM, a training needs assessment is necessary to identify the skills to be trained (Flin et al, 2008). To illustrate, extensive background research was carried out to identify the seven skills that form the basis of U.S. Naval aviation CRM training (decision making, assertiveness, mission analysis, communication, leadership, adaptability/ flexibility, and situational awareness; Prince & Salas, 1993). The needs assessment ensured that the training was based upon a sound research foundation and that limited time and resources for training were being effectively utilized.

G. CONCLUSION

To identify the applicable nontechnical skills for OODs, the author used a combination of literature review, focus groups, and interviews. Statements extracted from the interviews provided the tools that were necessary to test the reliability of the taxonomy when applied to actual surface Navy situations. Testing the inter-rater agreement multiple times shaped the list of skills into a much more effective and usable tool. The next chapter discusses how the author used the nontechnical skills taxonomy to create a rating form that could be used to assess the abilities of OODs.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY, RECOMMENDATIONS, AND CONCLUSION

A. SUMMARY

This thesis started with the goal of creating a behavioral marker system that could be used for assessing the nontechnical skills of OODs. The literature clearly indicates that human error is a leading cause a major mishaps, and the lack of nontechnical skills is a leading cause of human error. This situation is especially true in high-risk organizations such as the Surface Navy. Because it is impossible to create a system capable of addressing an entire ship, the OOD was the logical place to begin.

The author used a combination of literature review, focus groups, interviews, and inter-rater reliability testing to identify the nontechnical skills that are applicable to the OOD watch station. Once the skills were identified, the author used the taxonomy to create the Nontechnical Skills for Officers of the Deck (NTSOD) rating form. The NTSOD system can be used to observe and assess the nontechnical skills of OODs and candidates trying to complete the qualification. The author believes that with consistent use, the NTSOD system can increase the effectiveness and safety of OODs.

B. RECOMMENDATIONS FOR FURTHER RESEARCH

1. Recommendation 1

Additional research should be conducted to test the NTSOD system's reliability and usability while being used in the fleet or simulators. This research could support the findings that the NTSOD system accurately assesses the nontechnical skills of OODs. The research could also investigate the effects that the NTSOD system has on the overall performance of OODs.

2. Recommendation 2

The results of the present study should be used to provide a research foundation to Bridge Resource Management (BRM) training so that it more effectively addresses the skills required to safely stand the watch.

3. Recommendation 3

While the OOD is the most important watch station onboard surface ships, there are many other important positions such as Tactical Action Officer (TAO) and Engineering Officer of the Watch (EOOW). Similar watches also exist on submarines and merchant marine shipping. The methods described in this thesis should be used to develop behavioral marker systems for addressing the nontechnical skills of those watch stations as well.

C. CONCLUSION

Hetherington et al (2006) stated that there are many gaps in the maritime human factors literature. The NTSOD taxonomy is a step towards filling these gaps. Although further research is required to establish the reliability and validity of the NTSOD taxonomy, it is one of the first research-based behavioral marker systems developed for use in a maritime environment.

Human error can never be eliminated. However, identifying, training, and giving feedback on the nontechnical skills that are required for safe and effective performance will ensure that sailors have the appropriate skills for minimizing, detecting, and mitigating error before it leads to an accident.

LIST OF REFERENCES

Booher, H. R. (Ed.). (2003). *Handbook of human systems integration.* John Wiley & Sons, Inc.

Carter-Trahan, A. C. (2009). *An examination of the human factors attitudes and knowledge of surface warfare officers.* Master's thesis. Monterey, CA: Naval Postgraduate School.

Department of the Navy. (2005). *OPNAV instruction 3120.32C, change transmittal 6.* Washington, DC: Office of the Chief of Naval Operations.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4).

Fletcher, G., Flin R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2004). Rating non-technical skills: developing a behavioral marker system for use in anaesthesia. *Cognitive Technical Work, 6,* 165–171.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: A guide to non-technical skills.* Burlington, VT: Ashgate Publishing Company.

Flin, R., Yule, S., Paterson-Brown, S., Rowley, D., & Maran, N., (2006). *The non-technical skills for surgeons (NOTSS) system handbook v1.2.* University of Aberdeen: Central Printing Service.

Gatfield, D. (2005). Using simulation to determine a framework for the objective assessment of competence in maritime crisis management. *SAGSET 2005 Conference.*

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37,* 401–411.

Jones, D., & Endsley, M. (1996). Sources of situation awareness errors in aviation. *Aviation, Space, and Environmental Medicine, 67*(6), 507–512.

Kanki, B. G., Helmreich, R. L., & Anca, J. (2010). *Crew resource management: Second edition.* San Diego, CA: Academic Press.

Kirwan, B., & Ainsworth, L. K. (Eds.). (1992). *A guide to task analysis.* Bristol, PA: Taylor & Francis Inc.

Klampfer, B. Flin, R., Helmreich, R. L., Häusler, B., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, P., & Amacher A. (2001). Enhancing performance in high risk environments: Recommendations for the use of behavioral markers. *Behavioral Markers Workshop. Zurich, Switzerland*.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*, 456–460.

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1)*, 159–174.

Long, W., O'Connor, P., & McCauley, M. (2010). *Nontechnical skills for officers of the deck: A user's guide*. Monterey, CA: Naval Postgraduate School.

Miller, N. L., & Shattuck, L. G. (2004). A process model of situated cognition in military command and control. *2004 Command and Control Research and Technology Symposium.*

National Transportation Safety Board. (1988). *Marine accident report: Ramming of the Spanish bulk carrier Urduliz by the USS Dwight D. Eisenhower (CVN 69) Hampton Roads, Virginia August 29, 1988.* Washington, DC: National Transportation Safety Board.

O'Connor, P., & Cohn, J. (2010). Enhancing human performance in high reliability organizations: learning from the military. In: P. O'Connor & J. Cohn (Eds.). *Human Performance Enhancements in High-Risk Environments: Insights Developments, and Future Directions from Military Research* (pp. 1–8). Santa Barbara, CA: ABC- Clio.

O'Connor, P. & O'Dea, A. (2007). The U.S. Navy's aviation safety program: a critical review. *International Journal of Applied Aviation Studies, 7(2),* 312–328.

O'Connor, P., O'Dea, A., Flin, R., & Belton, S. (2008). Identifying the team skills required by nuclear power plant operations personnel. *International Journal of Industrial Ergonomics, 38,* 1028–1037.

O'Connor, P., O'Dea, A., & Melton, J. (2007). A methodology for identifying human error in U.S. Navy diving accidents. *Human Factors, 49*(2), 214–226.

Reason, J. T. (1990). *Human error.* New York: Cambridge University Press.

Reason, J. T. (1997). *Managing the risks of organizational accidents.* Brookfield, VT: Ashgate Publishing Company.

Roberts, K. H., & Rousseau, D. M. (1989). Research in nearly failure-free, high-reliability organizations: having the bubble. *IEEE Transactions on Engineering Management*, *36*(2), 132–139.

Rothblum, A. M. (2000). *Human error and marine safety*. U.S. Coast Guard Research & Development Center.

Secretary of Defense. (1988). *Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988.* Investigation Report. Washington, DC: The Secretary of Defense.

Sneddon, A., Mearns, K., & Flin, R. (2006). Safety and situation awareness: “Keeping the bubble” in offshore drilling crews. *SPE International Health, Safety, and Environment Conference, Abu Dhabi, U.A.E.*

Stavridis, J., & Girrier, R. (2007). *Watch officer’s guide: Fifteenth edition.* Annapolis, MD: Naval Institute Press.

Wagenaar, W. & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies*, *27*, 587–598.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons’ non-technical skills. *Medical Education*, *40,* 1098–1104.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Dudley Knox Library (Code 013)
   Naval Postgraduate School
   Monterey, CA

2. Defense Technical Information Center
   Ft. Belvoir, VA

3. W. Max Long
   Operations Research Department
   Naval Postgraduate School
   Monterey, CA

4. Paul E. O'Connor
   Centre for Intervention and Structural Change
   National University of Ireland
   Galway, Ireland

5. Michael E. McCauley
   Operations Research Department
   Naval Postgraduate School
   Monterey, CA

6. Kyle H. Turner
   SWOSCOLCOM
   Cushing Rd.
   Newport, RI